DOCUMENT RESUME

ED 170 385 TM 008 711

AUTHOR Daillak, Richard; And Others
TITLE Studies of School Evaluation Practice.
PUB DATE 31 Mar 78
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Decision Making; Educational Practice; Elementary Secondary Education; Evaluation Needs; Evaluators; Experimental Programs; *Information Utilization; Program Administration; *Program Evaluation; Research Utilization
IDENTIFIERS Elementary Secondary Education Act Title IVC

ABSTRACT

To examine the impact of program evaluation information upon decision making, case studies were conducted of five different programs: (1) a diagnostic teaching kindergarten program (K-LEAD), funded under the Elementary Secondary Education Act (ESEA) Title I; (2) Project MORE, an ESEA Title IV-C funded junior high school program dealing with school failure, discipline, and truancy; (3) an individually guided education program designed to improve achievement and morale; (4) a career education program--orientation toward the world of work--a Title IV-C project involving an entire high school; and (5) an innovative bilingual education program using the language experience approach at the elementary school level. Interviews were conducted with key staff who were involved in each of these program evaluations. The results indicated that the evaluation data tended to have influence, but not absolute impact, upon the decisions made by local program staff. Since program directors were typically the key decision-making audience for the evaluations, their relationships with the evaluators were very important. It was suggested that evaluators should provide information that is useful to program directors, in order to encourage utlization of the evaluation results. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Mark Young
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

STUDIES OF SCHOOL EVALUATION PRACTICE*

Richard Daillak, Marvin C. Alkin and Peter White
CENTER FOR THE STUDY OF EVALUATION, UCLA

ED170345

Carol Weiss has asserted that:

> The basic rationale for evaluation is that it provides information for action. Its primary justification is that it contributes to the rationalization of decision-making. Although it can serve such other functions as knowledge-building and theory-testing, unless it gains serious hearing when program decisions are made, it fails in its major purpose. (Weiss, 1966)

Measuring evaluation by this standard, there is some question whether educational evaluation is succeeding in its basic purpose. As has been noted in a recent study of evaluation utilization, "there seems to be a consensus that the impact of evaluative research on program decision-making has been less than substantial." (Patton et al, 1975)

The dissatisfaction with evaluation's lack of usefulness and impact has been so great that quite an extensive literature has grown up around the problem of evaluation under-utilization, beginning with a wave of articles in the mid-sixties (e.g. Rossi, 1966; Dexter, 1966) and extending to the present (Patton, in press). Unfortunately, though, this literature has remained for the most part at the level of anecdote and speculative explanation; only a handful of the many articles dealing with evaluation utilization draw upon any empirical research base.

The few empirical studies that have been conducted, however, seem to present a picture which is less gloomy than many of the more speculative articles; they suggest that evaluation can have impact upon decision-making, although not necessarily the kind of dramatic go/no go influence some would wish. For example, in one of the earliest empirical studies, Alkin and his associates (1974) at the Center for the Study of Evaluation examined the impact of evaluation on decision-making in forty-two ESEA Title VII programs. The findings of this study show that the influence of evaluation on federal level decision-making was, as the literature would predict, minimal. But quite different findings emerged regarding evaluation impact at the local level. Here the researchers note:

TM008711

*Paper presented at the meeting of the American Educational Research Association, March 31, 1978, Toronto, Canada. The activity which is the subject of this report was supported in whole or in part by the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

> With respect to Title VII evaluations, the predictions of disuse of evaluative information were not fulfilled. The Project Director Questionnaires indicated that evaluation activities were very useful to project directors. Only one project director reported that the final (year end) evaluation report was "not at all" influential in decisions to modify the project the following year....Sixty-four percent "agreed" or "strongly agreed" that evaluation influenced decisions that were made to modify the project during the year.... When asked if the evaluation had been of assistance to them,project directors strongly contended that it had been. This applied in many areas but particulaly in "identifying possible problems," "preparing reports," and "providing general recommendations for program changes." (p. 87)

This same sort of finding was even more clearly demonstrated in a subsequent study by Michael Patton and others (1975) at the Minnesota Center for Sociological Research. These researchers looked at twenty evaluation studies of national health programs and, for each program studied, interviewed the evaluator and a major program decision-maker. The researchers stated:

> What we found in response to these question on impact was considerably more complex and less dismal than our original thinking had led us to expect. We found that evaluation research is used by decision-makers but not in the clear-cut and organization-shaking ways that social scientists sometimes believe research should be used... The results of our interviews suggest that what is typically characterized as under-utilization or non-utilization of evaluation research can be attributed in substantial degree to a definition of utilization that is too narrow and fails to take into consideration the nature of actual decision-making processes in most programs. (p. 10)

Thus, in the Alkin study, which looked at evaluations being conducted at the local level, and in the Patton study of evaluations national in scope, we see evaluation being used successfully for on-going program development and management, and having incremental rather than dramatic impact upon program decision-making. These studies suggest that the key to understanding evaluation utilization (and to increasing utilization) may lie in the careful study of the real world of the evaluation practitioner and user. In schools and school districts everywhere, there are knowledgeable educators trying to conduct good evaluations and trying to make intelligent use of evaluation data. Operating, as they must, within problematic settings, they are still able to devise strategies to deal with the obstacles to good evaluation. By identifying the strategies they have employed and by contrasting and comparing these strategies, successful and unsuccessful, we may learn a great deal about making evaluation more useful.

The research reported on briefly in this article, and at greater length elsewhere (Alkin et al, in press), is an initial step in this direction. Our focus was the degree and manner of evaluation impact on school programs, and our approach was to examine carefully, but with a minimum of preconceptions, the process by which several program evaluations were carried out and by which the level of impact was determined.

Methodology - The Intensive Case Study

Specifically, we decided to conduct a series of case studies of evaluations and their utilization. Each study would focus on a different elementary or secondary school program and its evaluation history and would describe the program, the program staff, the program's evaluation (or perhaps several consecutive annual evaluations), and the evaluation staff. The objectives of the study would be to put together a complete, and accurate, picture of the evaluation process within the program--looking particularly at those persons who shaped that process, how the evaluation fit into the total operation of the school program, and in what way the evaluation influenced decisions made about the program.

A retrospective interview approach to the case studies was selected. This approach involved interviewing, in depth, the operational staff and the evaluator of an educational program which had been selected for study, and, as a supplement to the interviews, reviewing documentary evidence such as program proposals, evaluation reports, and the like. The programs selected for study were all funded through ESEA, all were in at least their second year of operation, and, because the programs were evaluated annually, each had gone through at least one full evaluation cycle.

In locating potential sites, we capitalized on a variety of resources. Personal and professional contacts in public educational agencies were especially helpful in directing our attention to potential sites. The approachability of the site personnel was a clear consideration since we did not have the power, or the inclination, to impose ourselves on the case study hosts. As a result, we cannot claim that the five chosen sites were randomly selected.

Initial contact was established either with the site program evaluator, program director, or school district administrator, and arrangements were made for an exploratory meeting between ourselves and program personnel. A common outcome of the initial meeting was for us to obtain copies of significant program documents for later review. These documents--funding proposals and progress reports, evaluation reports, and so forth--allowed us to familiarize ourselves with at least the formal descriptions of the program and its evaluation, and provided a helpful background for the next step of the case studies: the data collection interviews.

A first round of private interviews was scheduled with the site program director, the program evaluator, the Principal, and any other persons intimately associated with program decision-making. In this first round of interviews, every attempt was made to avoid being overly directive--to allow the interviewees, the "informants," to tell the story of the program and its evaluation in their own words, letting each person choose the topics they

thought important to discuss. Our concern was to avoid imposing our own a priori biases on the conversation, biases which could lead the informants into telling us the story as we might want to hear it rather than as they would naturally present it.

At the same time, our reading of the literature did alert us to at least look for the operation of a variety of factors: the social and organizational context of the program; the characteristics of the evaluator, program staff and program administrators, the purposes and procedures of the evaluation, the kinds of interactions occurring among the major actors, and the manner in which evaluation findings were communicated. Usually, we simply indicated our broad interests--the history, context, and events of the program and its evaluation, and the impact of the evaluation on the program--and then let the informant guide the conversation for most of the remainder of the interview, asking occasional clarifying questions as necessary. Toward the end of the interview period, the interviewer would try to hit on topics or issues that had not already been raised but which appeared to be worth exploring. In fact, the commonsense accounts given by the informants touched on most of the broad topics listed above, so the task of the interviewer was considerably eased.

Significantly, the interviews with the core informants (the evaluator, the program director, and perhaps one or two others) frequently led us to other potential interviewees--for example, a particularly influential teacher, or principal, or counselor. These newly identified informants, when interviewed, might in turn lead us to other persons; thus, the interviews had a sort of snowball effect until all the major actors in the evaluation situation had been identified and, if possible, interviewed. After this point, when all the major actors had seen interviewed and when any follow-up interviews (for clarification, for example) had been completed, we were ready to begin drafting the case study reports.

In writing the reports, we did, inescapably, exercise our own judgment in fitting together the pieces of the story, and there is always the possibility that we may have unintentionally misread the data and, hence, misrepresented the facts. There is, in addition, the possibility that some of the informants may have biased our interpretation in favor of their point of view. We did, however, employ a number of safeguards against these possibilities. Perhaps the most important of these were the procedures employed in the drafting and revision of the case study reports.

As a first, rough, screen against error, the initial draft of the case study was reviewed by members of the research team itself. Thus, the draft was revised to satisfy the research team rather than a single author. Unlike many field research reports, the process of review did not end with the research team. The case study draft was next circulated to the interviewees themselves for comment and criticism. Each key informant was given a copy of the case study report to examine, and a personal interview was scheduled to discuss the report.

The informants did not always agree with the draft reports, as might be imagined. In the personal interviews, informants were encouraged to state their opinions, and objections. Frankly, in most cases, the modifications suggested by the informants were readily accepted. We were fallible, and we had occasionally made factual errors or misinterpreted an informant's statements. Some criticisms were more far-reaching, touching upon the overall tone of the case study report or the research team's analyses of the evaluation process. Each such criticism was explored in depth with the critic, and, when necessary, telephone conversations of personal interviews were conducted with other informants in order to thrash out the issue.

Based upon the additional information gained in these review sessions, the case study report was edited into final form by the research team. All in all, the document was what we felt to be fair and accurate in light of all the data.

As a final effort at fairness, our key informants were invited to submit remarks commenting upon the case study findings, the case study process, and the study of evaluation and evaluation utilization. Their remarks will be included along with the case study reports in a forthcoming full-length report of the research. (Alkin et al, in press)

The Cases

In this section, we will present a paragraph or two describing the nature of each evaluated program. It is difficult-- indeed impossible -- to give the real flavor of the case study reports in this short paper. Thus, we have chosen not to attempt to collapse the case studies but to provide only a quick introduction to each case situation. The interested reader is referred to the full-length report of the research for the case study descriptions.

Rockland

This case study focused on the evaluation of K-LEAD, a kindergarten program in the Rockland school district funded under ESEA Title I. K-LEAD, a program emphasizing the diagnosis of the child's learning needs and the prescription of individualized treatments to meet those needs, was the cornerstone of the Rockland district's Title I effort. The approach employed in K-LEAD had a five-year history in the district, and K-LEAD itself was an established program which had earned a reputation among teachers and administrators as successful and well-run.

The evaluation of K-LEAD was envisaged as a critical test of that reputation. As Mrs. Anne Stuart, the local Title I program director noted, the reputation of K-LEAD was based largely on subjective data, and the time had arrived for a more precise, objective look at the educational costs and benefits of K-LEAD. A central question, then, was: Was K-LEAD really better than the less intensive, standard kindergarten approach?

A second focus of the evaluation was more limited, involving the pilot testing and evaluation of a possible supplement to K-LEAD. The Norton Music Program, a commercially available course, had been suggested as a potential aid to K-LEAD students in the development of auditory skills. The Norton Program, however, was rather costly, and Mrs. Stuart felt that the expense could only be justified if Norton produced substantial benefits over and above the existing K-LEAD program.

The evaluation findings indicated that the K-LEAD program was significantly more effective than the regular kindergarten program; however, it also indicated that the addition of the Norton supplement did not improve student performance over and above the usual K-LEAD level. The decision was made by the program director to continue K-LEAD but to drop plans for using the Norton Music Program as a supplement.

Bayview

The evaluation of Project MORE in Bayview followed quite a different path from that of the Rockland case. Unlike Rockland's K-LEAD, Project MORE was a new experimental (ESEA Title IV-C) program restricted to one school in the Bayview district (McNaught Junior High), and the case study focuses on the first and second year evaluations of this developing program.

Project MORE, operating in a lower and lower-middle class area, sought to ameliorate the high rate of school failure and the worrisome discipline and truancy problems which McNaught administrators and staff perceived as a major problem. Central to the strategy adopted for Project MORE was the belief that school failure, behavior problems, and truancy were interlinked; time lost from school through behavior related suspensions and truancy contributed to poor attitude toward school, behavior problems, and truancy--and so on in a vicious downward spiral. Project MORE sought to interrupt that spiral and reverse its direction.

The evaluation produced no earth-shattering changes in Project MORE, but it did have an influence on classroom practices, and the evaluator was able to "sell" the initially skeptical project staff on the notion that evaluation could be relevant to them. The end result was that evaluation-produced information was consulted by project decision-makers, even though it remained but one decision-making input among many.

Valley Vista

This case study focused on the evaluation of the Wisconsin IGE (Individually Guided Education) package adopted by the Sierra Elementary School. Sierra served a predominately

poor and middle class population and experienced a high rate of student transiency and a chronic shortage of classroom space. Student test scores at the school ran below national averages and were among the lowest in the Valley Vista district.

Faculty morale at Sierra was dampened by these conditions, and the level of frustration was high. The Wisconsin IGE program, with its emphasis on teacher involvement and autonomy and its claims of significant student gains in achievement, was an attractive new possibility to administrators and staff looking for some sort of solution to the school's problems. Using Title I funds, IGE was implemented in Sierra in the fall of 1973.

The evaluation of IGE at Sierra focused on the collection of achievement data from standardized tests and the monitoring of compliance with program plans. Although there was modest use of the achievement test data -- for individual pupil assessment and the identification of general program strengths and weaknesses -- there was also some feeling that the evaluation was overly burdensome and only marginally relevant.

Clayburne

Project OUTREACH, and ESEA Title IV-C project in the Clayburne district, consisted of an entire high school focused on "career education." Unlike vocational training, which emphasizes training for some specific job, career education is a more comprehensive, integrated approach to the "world of work." The school seeks to develop in students knowledge of, and positive attitudes toward, the world of work and the ability to explore and exploit a variety of job opportunities. Academic programs are designed to complement and integrate with this career education, and the curriculum is organized around career clusters -- e.g. business, construction, manufacturing, urban agriculture -- so that a student selecting a concentration in one or the other of the clusters learns a set of skills common to the variety of occupations within that broad cluster. This gives the student greater flexibility than the narrow job training of some more traditional programs.

The most striking feature of the Clayburne case study was the intensive involvement of the evaluator, Dr. Lamar Gerry, in the broad range of program concerns throughout the history of the project. Dr. Gerry's expertise -- his professional training was in vocational education was a valuable resource for the decision-makers of Project OUTREACH as they grappled with the many problems facing the new program.

Garrison

This case study examined the evaluation of an innovative bilingual education program at the McNair Avenue Elementary School in Garrison. The program employed a "language experience" approach to the teaching of reading, the basic premise of which is that the

child can learn to read what he or she can say. Under this approach, instruction builds upon the existing linguistic competencies of the child and is individualized so that the child defines the basic content of the reading lesson. Writing instruction is similarly grounded in the experiences of the child and centers around stories and poems with themes relevant to the daily lives of the children. This approach is especially exciting in a bilingual-bicultural program because the child - authored pieces can be recycled into the program as instructional materials.

The introduction of the program to McNair generated some controversy in the community, especially among parents concerned that the program might sacrifice competency in English in favor of Spanish language competency. Furthermore, the program was funded under ESEA Title IV-C, whose SEA overseers place great emphasis on test scores and cognitive achievement. Under these circumstances, the evaluation focused on documenting student achievement through the use of standardized tests, and a contract for the evaluation was given to an external evaluator with both fine academic credentials and also political stature and credibility within the barrio community.

Thus, from the beginning, political and other external factors were important in the McNair situation. The evaluation was embedded in this context, and the story of the evaluation and the use made of its findings -- which were very credibly positive -- makes fascinating reading.

Discussion

Without going into the details of each study, it is possible to sketch out a number of commonalities or thrusts of the research findings. One clear observation is that none of the evaluations we studied had make or break impact upon the programs they assessed. Perhaps the closest any came to such dramatic impact were the third year evaluations of the Title IV-C funded programs (in Bayview, Clayburne, and Garrison). The third year evaluations fairly closely scrutinized by the SEA overseers of Title IV funds as they consider whether or not to grant final "dissemination funding" to the program for the next year. Yet, the reality is that very few programs are granted the dissemination funds, so most local program people expect a negative decision and plan around it. All three of the Title IV programs we examined had their applications for fourth year dissemination funding turned down, yet each program was able to find the funds to continue, albeit with occasional program modifications or reductions.

Major modifications in the programs as a direct result of the evaluations were also quite rare. What was more common was gradual, incremental impact. In Bayview, there was influence on teacher classroom behavior; in Clayburne, the evaluator made a variety of recommendations about curriculum, student selection, and the like, which were followed up by the principal and teachers; while in Valley Vista and Garrison, the test score data and other evaluation findings were looked at by local program decision-makers and used -- very modestly -- to identify program strengths and weaknesses. It was only in Rockland, in the decision not to use the Norton music component, that we found a major program decision being made almost completely on the basis of evaluation data.

So, the typical pattern was for evaluation data to have influence, not decisive impact, upon the decisions made by local program staff. We found little evidence that evaluation had significant impact on decisions made about the programs by Federal or SEA decision-makers (this supports an earlier finding by Alkin et al, 1974, in a study of Title VII evaluations), and the responsible school district-office administrators typically did not interfere with program operations. Thus, program directors were usually the key decision-making audience for the evaluations.

Under these circumstances, the relationship between the evaluator and the program director assumes considerable importance. The evaluator who focuses the evaluation on the needs of the program director seems more likely to create an influential evaluation. This may involve "selling" an initially skeptical director on the value and relevance of evaluation, emphasizing formative concerns more than summative assessment, and minimizing the burden the evaluation places upon the program. Of course, there is a tension between serving the program director in this fashion and fulfilling the mandated evaluation role, which centers on summative

evaluation, compliance, and accountability. The evaluator who concentrates on the mandated evaluation tasks runs the very real risk of losing the local audience; the most successful evaluators we met dutifully carried out the required tasks of their work -- collecting and analyzing standardized test data, filling out the required reporting forms, etc. -- but tended to invest greater energy and creative effort in devising ways to help local program people with the small, day to day decisions that cumulatively make a big difference.

Situations vary, of course, and there are times when the more traditional notions of evaluation are exactly what local program people need and want. The Rockland experimental tryout of the Norton music program was an example of this, and in the Garrison case the school principal wanted the evaluator to focus on student achievement so that there would be "proof-positive" that the bilingual program was effective (proof which the principal then used politically and organizationally to combat the critics of the program and to secure additional funding).

In conclusion, then, we have observed a number of instances of evaluation which have had influence upon program decision-making. Tailoring the evaluation to the needs of the local programs -- and sometimes educating local decision-makers in the ways evaluation can be relevant to their needs -- can be important aids to utilization. Recognizing the limitations of evaluation is also important. The evaluator can advise decision-makers and serve as a trusted consultant, but it is unrealistic, and perhaps presumptuous, to expect evaluation to command the situation. The evaluator who renounces the role of omniscient overseer of the program in favor of the more humble role of advisor may, in the long run, be more effective in achieving the ultimate goals of evaluation: informing action and promoting rational decision-making.

References

Alkin, M.C., Daillak, R.H., & White, P.S. Evaluation Utilization in Local Schools: A Case Study Approach. Beverly Hills, California: Sage Publications, in press.

Alkin, M.C., Kosecoff, J., Fitz-Gibbon, C., & Seligman, R. Evaluation and Decision-Making: The Title VII Experience. CSE Monograph Series in Evaluation, (No. 4). Los Angeles: Center for the Study of Evaluation, UCLA, 1974.

Dexter, L.A. Impressions about Utility and Wastefulness in Applied Social Science Research. American Behavioral Scientist, 1966, 9 (6), 9-10.

Patton, M.Q. Utilization Focused Evaluation. (Sage Library of Social Research, Vol. 65) Beverly Hills: Sage, in press.

Patton, M.Q., Grimes, P.S., Guthrie, K.M., Brennan, N.J., French, B.D., & Blyth, D.A. In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research. Minneapolis: Minnesota Center for Social Research, University of of Minnesota, 1975.

Rossi, P.H. Boobytraps and Pitfalls in the Evaluation of Social Action Programs. Proceedings of the Social Statistics Section. Washington, D.C.: American Statistical Association, 1966, 127-132.

Weiss, C.H. Utilization of Evaluation: Toward Comparative Study. Paper presented at the American Sociological Association Meeting, Miami Beach, Florida, September 1, 1966. Reprinted in C.H. Weiss (Ed.), Evaluating Action Programs. Boston: Allyn and Bacon, Inc., 1972.